# Statistical analysis of the sampling design: FishPi case study on the biological sampling of the European hake fishery

Jose Castro, Nuno Prista, Lucia Zarauz, Jose Rodriguez, Manuela Azevedo, Laurent Dubroca, Alastair Pout, Joao Pereira, Chun Chen, Hans Gerritsen, Jon Elson, Ana Ribeiro, Patrik Börjesson, Kirsten Birch Håkansson, Sofie Nimmegeers

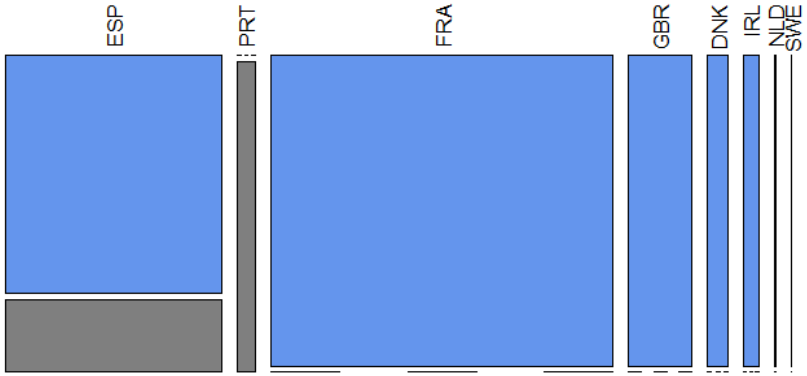
WP2 / The fishPi project
Regional co-operation in fisheries data collection

This work package was tasked with running **simulation models of the sampling** to test regional sampling designs for the fisheries exploiting regional stocks.

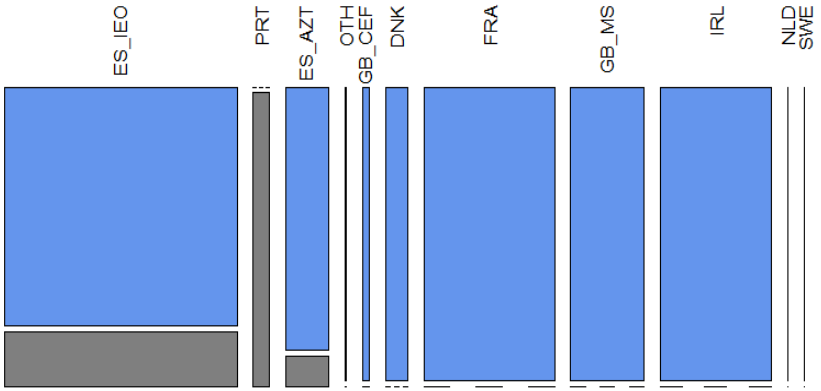
This simulation studies were prepared to **mimic the random selection of fishing trips** from the regional population of trips for that fishery.

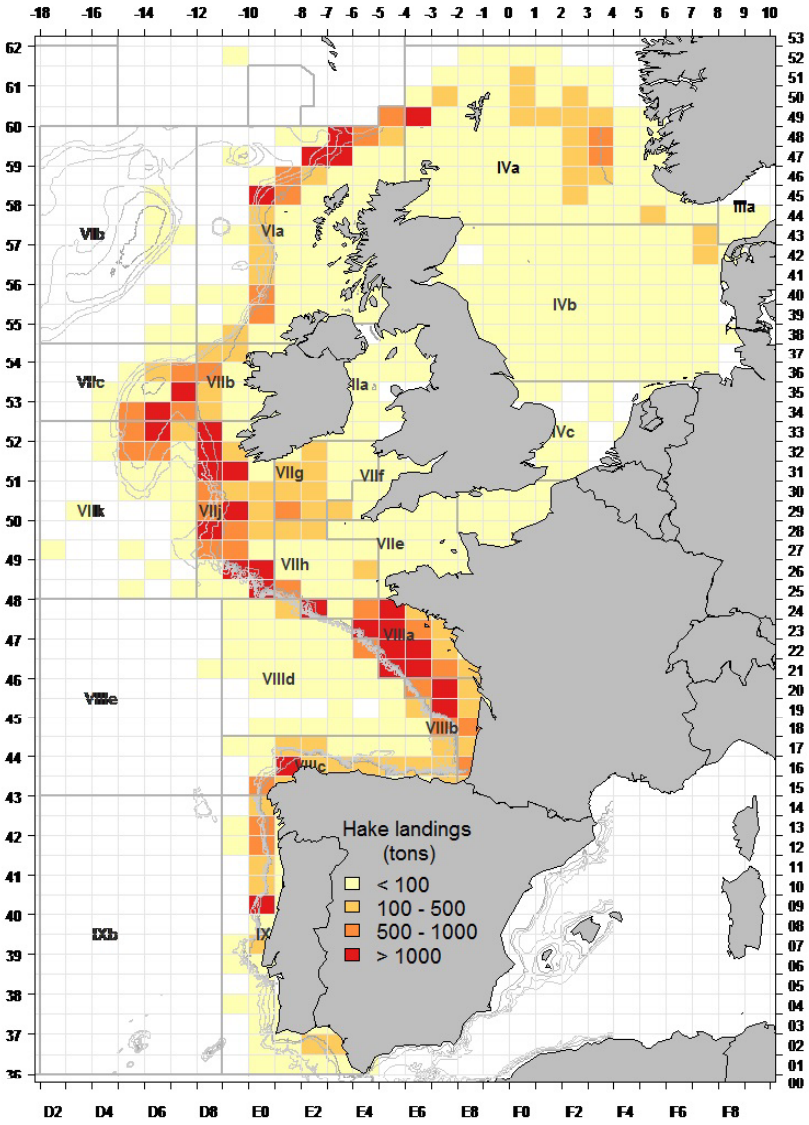
The study focus on the statistical optimization of **landings estimates**.

## Landings (mean 2013-2014) by vessel flag




## Landings (mean 2013-2014) by sampling Institute




Field "**onShoreSampLoc**" indicates the **country responsible for the sampling** and is fundamental to organize a regional sampling program.

## CS4-total: list of scenarios

2-stage random sampling
PSU: market/port and day
SSU: trip

1. Baseline scenario: SRS
2. Scenario COUNTRY
3. Scenario INSTITUTE
4. Scenario REGIONAL stratified by PORT
5. Scenario COUNTRY and PORT
6. Scenario INSTITUTE and PORT

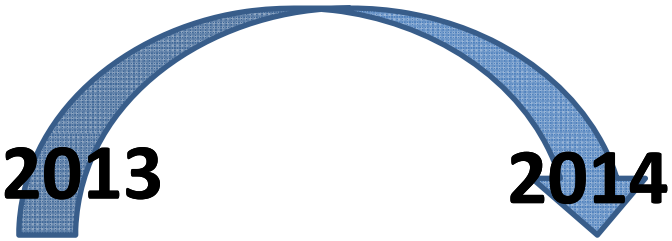
## CS4-southern: list of scenarios

2-stage random sampling
PSU: market/port and day
SSU: trip

1. Base line scenario: SRS

2. Scenarios REGIONAL.
   - 2a REGIONAL stratified by port
   - 2b REGIONAL stratified by quarter
   - 2c REGIONAL stratified by port-quarter

3. Scenarios COUNTRY
   - 3a COUNTRY stratified by port
   - 3b COUNTRY stratified by quarter
   - 3c COUNTRY stratified by port-quarter




to set stratification
and sample allocation

to run simulations
of each scenario

## Main criteria used in selecting the ”best” scenarios

- Relative Bias:

  Ratio between the estimatesd values (tons of hake) and the true value registered in 2014

- Relative Precison:

  Ratio between the standard error (SE) of the estimated values of each scenario and the SE of Scenario 1 (*i.e.* 2-stage SRS not stratified)

- Coverage by country:

  n PSUs, n SSUs, n SSUs with hake, tons hake available for sampling

- Coverage by domain:

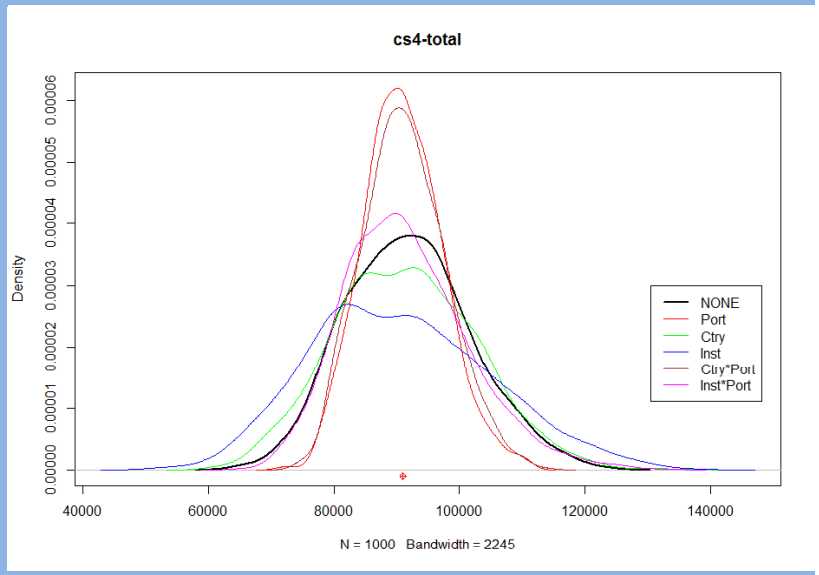
  n PSUs, n SSUs, n SSUs with hake, tons hake available for sampling

  - Country
  - Port and Quarter
  - Stock
  - FleetNat
  - FoCatEu6

## CS4 - TOTAL SIMULATIONS RESULTS

Highest precision
- scenario 4: **REGIONAL** STRATIFIED BY **PORT**
- scenario 5: **COUNTRY** AND **PORT**

| Scenario | Sampling design | Precision |
|---|---|---|
| 1 | 2-stage SRS (not stratified) | baseline |
| 2 | Stratified by Country | 1.14 |
| 3 | Stratified by Institute | 1.44 |
| 4 | REGIONAL stratified by PORT | **0.63** |
| 5 | Stratified by COUNTRY and PORT | **0.67** |
| 6 | Stratified by INSTITUTE and PORT | 1.00 |




| Institute | Current data | Port | Country/Port |
|---|---|---|---|
| | Nsamp_marketdays | SampTrips | SampTrips |
| ES_IEO | 765 | 1035 | 652 |
| PT | 700 | 109 | 700 |
| FR | 597 | 882 | 597 |
| GB_CEFAS | 431 | 38 | 132 |
| DK | 113 | 104 | 113 |
| IE | 82 | 226 | 82 |
| ES_AZTI | 43 | 230 | 156 |
| NL | 38 | 3 | 38 |
| GB_MS | 11 | 189 | 344 |
| SE | 3 | 1 | 3 |

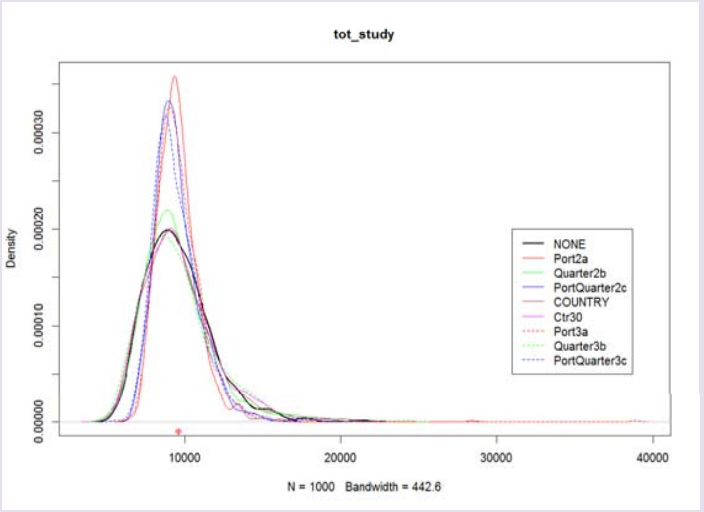
## CS4 - SOUTHERN SIMULATION RESULTS

Highest precision:

- scenario 2a: REGIONAL stratified by PORT
- scenario 2c: REGIONAL stratified by PORT-QUARTER

→ *The most precise overall but provide less coverage to some metiers and national fisheries.*

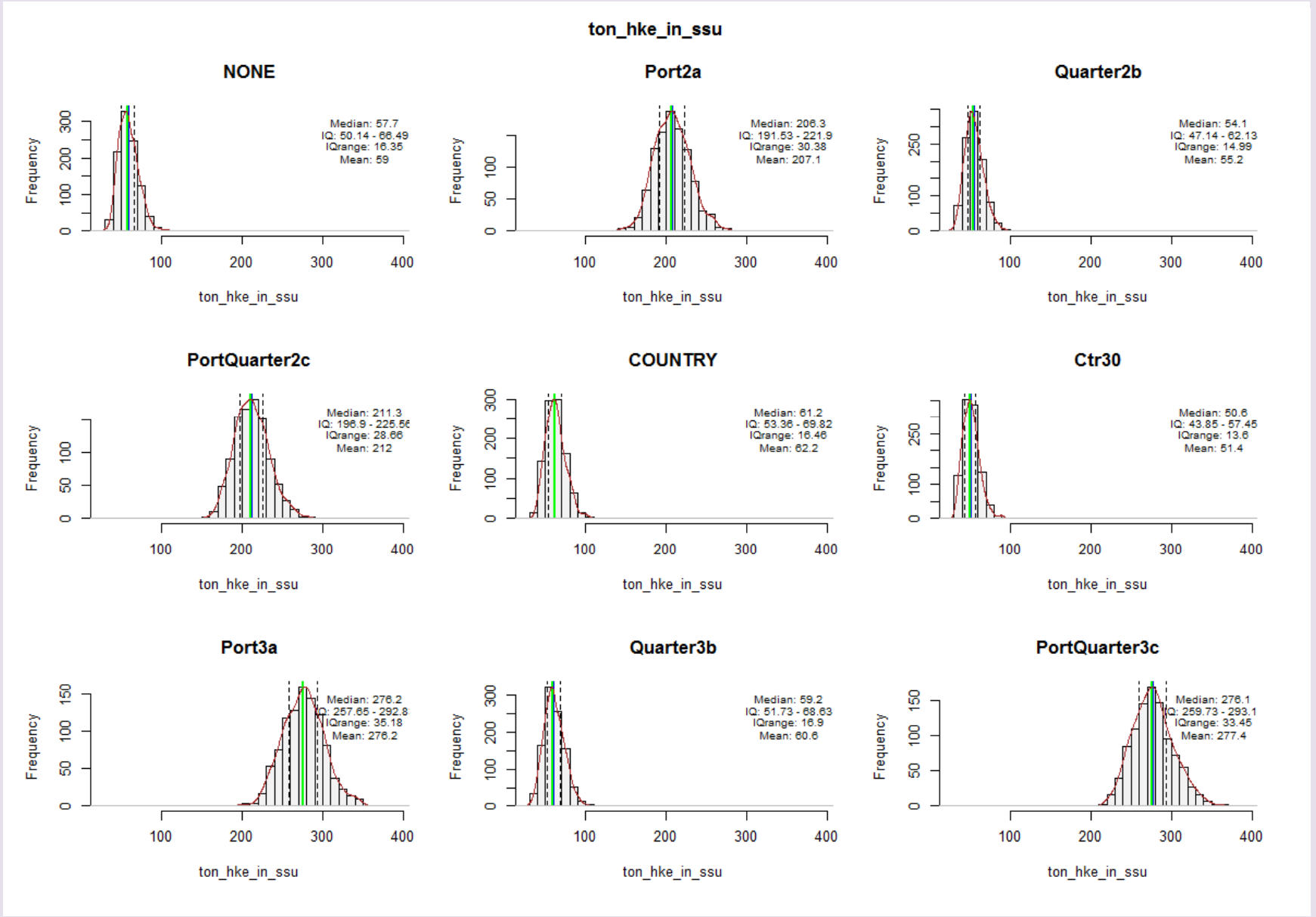
| Scenario | sampling design | Precision |
|---|---|---|
| 1 | 2-stage SRS WR (no stratified) | Baseline |
| 2a | Regional stratified by port | **0.67** |
| 2b | Regional stratified by quarter | 1.05 |
| 2c | Regional stratified by port*quarter | **0.68** |
| 3 | Stratified by country | 1.11 |
| 3a | Stratified by country and port | 0.80 |
| 3b | Stratified by country and quarter | 1.17 |
| 3c | Stratified by country and port*quarter | 0.76 |
| 3o | Stratified by country (proportional to landings) | 1.09 |



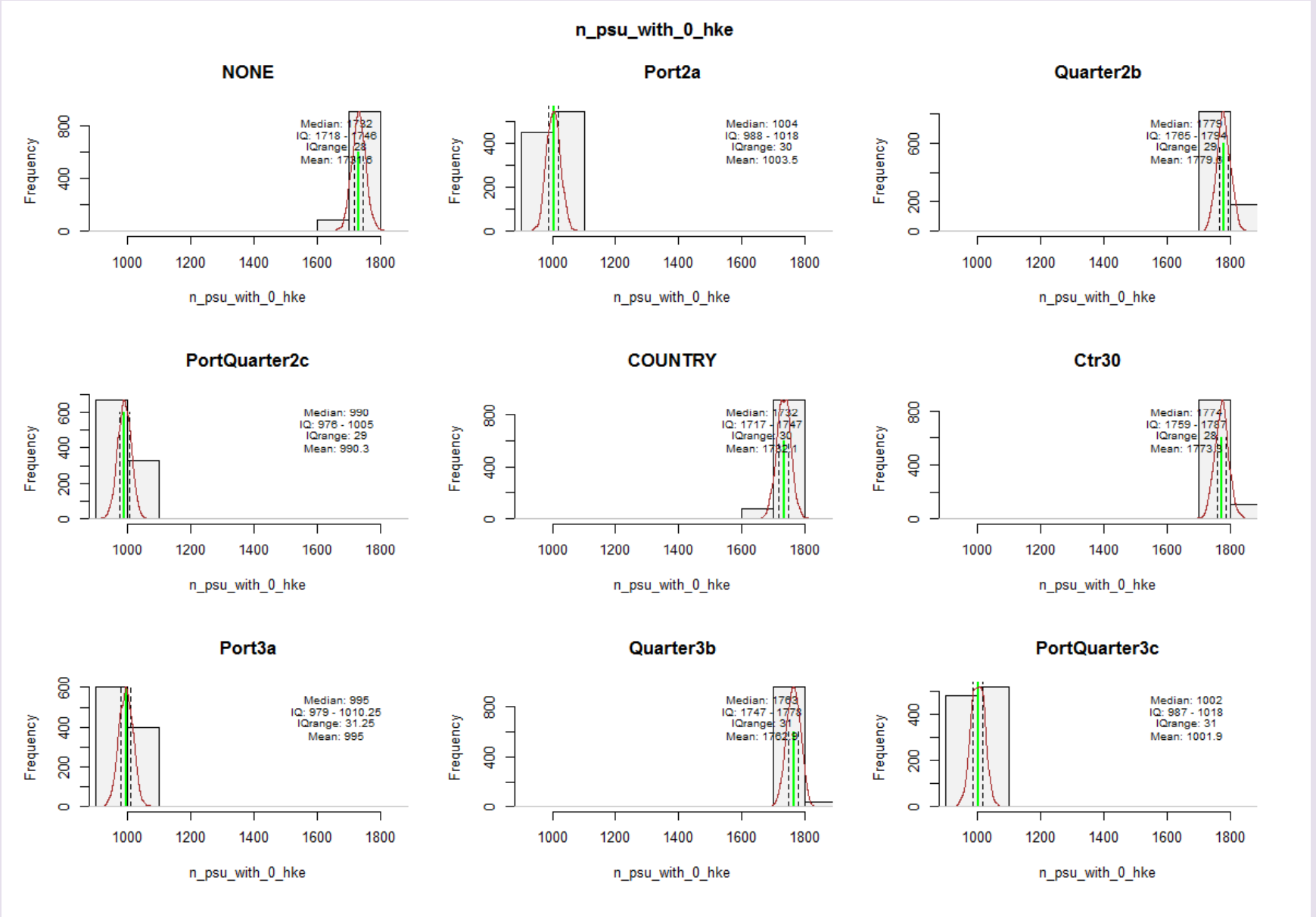

The second best-performing results were provided by scenarios 3a and 3c where the COUNTRY stratification is extended to PORT and PORT-QUARTER stratifications.

→ *Less precision in the estimation of total hake landings* ***but improved "tonnes of hake sampled"*** *and* ***"trips with hake sampled"*** *while providing* ***better coverage in some areas*** *with smaller hake landings.*

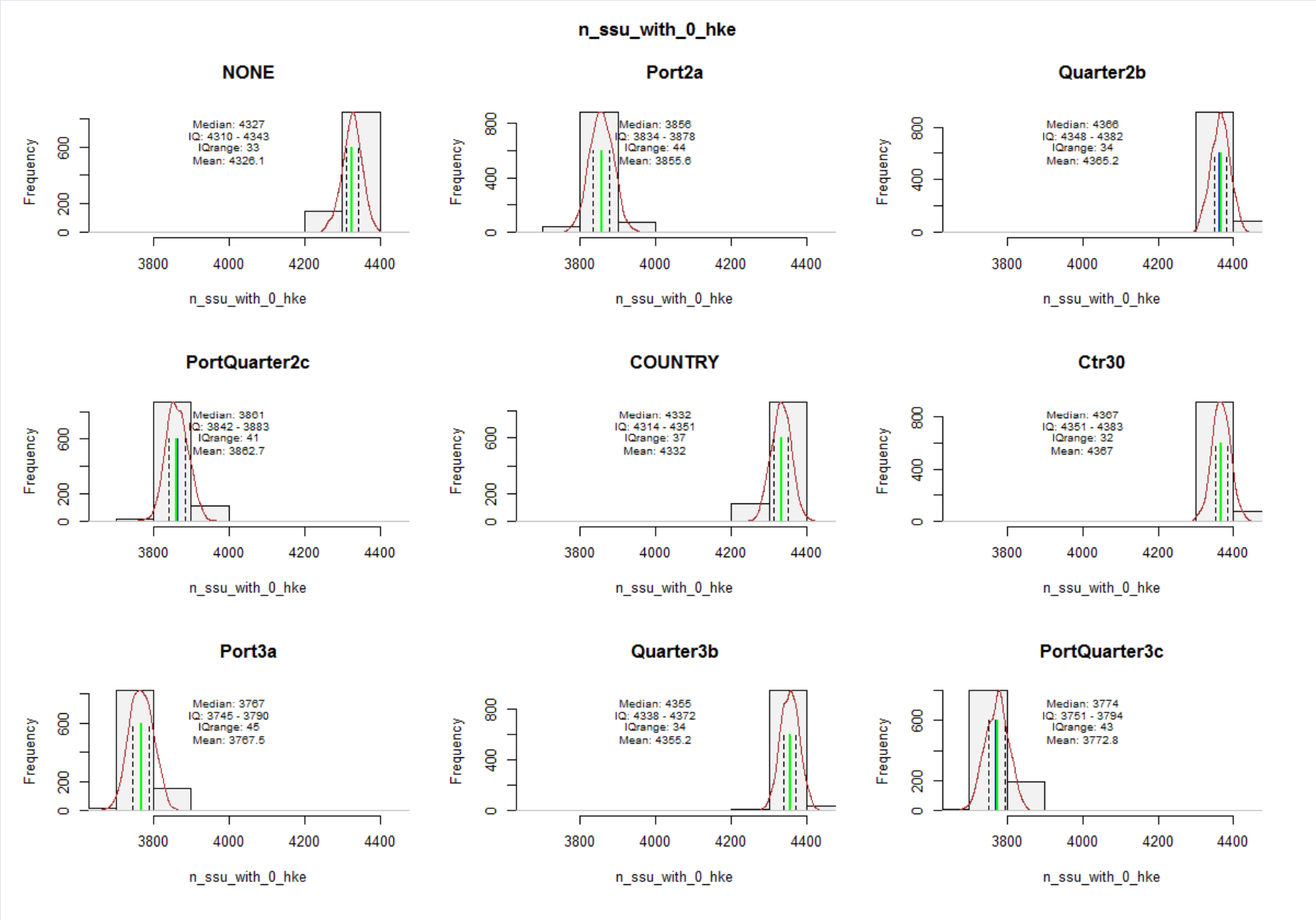
# CS4 - SOUTHERN SIMULATION RESULTS (examples: tonnes of Hke “sampled”)



# CS4 - SOUTHERN SIMULATION RESULTS (examples: market_days without hake)



# CS4 - SOUTHERN SIMULATION RESULTS (examples: trips without hake)



## CS4-southern hake: Methodological insights
## Comparison between scenarios with all days and positive trips

### Insights from CS4:

- Do we really know exactly in which market*days and trips our target stocks will occur?
- Does this affect our perceptions on the quality of designs and how much Regional Sampling Plans can deliver?

| **All market days All trips** | **All market days Positive trips** | **Positive market days Positive trips** |
|---|---|---|
| ↓ | ↓ | ↓ |
| All possible markets (and days) Random selection of market*days Random selection of **all trips** present in each markey*day | All possible markets (and days) Random selection market*days Random selection of **trips with HKE** in each | Only market (and days) that ***will*** register HKE Random selection of **trips with HKE** in each market*day |
| Design targets all fisheries (basic scenario for CS4 SthHKE) | Design targets hake fisheries (other species would be ”sampling by-catch”) | Design targets hake fisheries *realistic? can we assume to know the exact days of arrivals?* |

## CS4-southern hake: Methodological insights
## Comparison between scenarios with all days and positive trips

### Implications on size of the simulation dataset (2014)

| | all market*days all_trips | all market*days pos_trips | pos market*days pos_trips |
|---|---|---|---|
| No. Markets | 186 | 186 | 151 |
| Market*days with hake landings | 34849 | 34849 | 19074 |
| No. Market*days | 67890 | 67890 | 19074 |
| No. Trips | 584012 | 152465 | 103649 |

We ran simulations of these 3 setups using the same sampling effort and design Port2a (Regional Port Strata)

# CS4-southern hake: Methodological insights
## Comparison between scenarios with all days and positive trips

| | Expected value (mean) from 1000 sims | | | InterQuartile Range | | |
|---|---|---|---|---|---|---|
| type | all_marketdays all_trips | all_marketdays pos_trips | pos_marketdays pos_trips | all_marketdays all_trips | all_marketdays pos_trips | pos_marketdays pos_trips |
| tot_study | 9592 | 9638 | 9606 | 1553 → | 1068 → | 641 |
| ton_hke_in_ssu | 207 → | 503 → | 769 | 30 | 46 | 53 |
| | | | | | | |
| n_ssu_with_hke_27.8.c | 490 → | 1283 → | 2137 | 30 | 40 | 33 |
| n_ssu_with_hke_27.9.a | 472 → | 1531 → | 2687 | 29 | 44 | 33 |
| tons_hke_sampled_27.8.c | 160 → | 353 → | 542 | 29 | 43 | 49 |
| tons_hke_sampled_27.9.a | 47 → | 150 → | 227 | 8 | 15 | 17 |
| | | | | | | |
| n_ssu_with_hke_OTB_DEF_>=55_0_0 | 154 → | 473 → | 724 | 18 | 32 | 37 |
| tons_hke_sampled_OTB_DEF_>=55_0_0 | 29 → | 90 → | 131 | 9 | 15 | 20 |

→ Sharp increase in expectations

# Considerations for selections of regional design

**Datasets to use:** define strata and sampling intensity and simulate on different datasets; consider throughly if datasets with only positive trips are applicable

**Precision**: analyze the scenarios with acceptable precision.

**Coverage by country**: analyze the **deviations that occur with respect to the current coverage.**

**Coverage by domain**: analyze the **deviations that occur in estimates of domains currently used by the National institutes to provide commercial data to ICES**: technical (métier), spatial (ICES Division), and time (quarter) disaggregation.

**Indicators of efficiency:** consider indicators like mean and variance of *n_ssu_with_spp* and *ton_spp_in_sampled_ssu* per species, strata and post_strata) should be a part of outputs so we approximate better what we can expect from design.

**Match real-life fleet operation conditions**: E.g., all days of the year should be included unless specific days are known not to have any landings; all fleets should be included unless specific fleets are known to have negligible landings with similar fish size/age structures.

**Adequacy to multiple end-user needs**: A regional sampling design should aim at **estimating accurate length/age composition** of the landings of the hake stocks for stock assessment purposes while at the same time providing **accurate estimates for many other stocks and fisheries subject to assessment**. The current study focused on estimating total landings of hake which are hardly a proxy for size/age.